

# **Brain Tumor Detection Using Deep Neural Networks**



By

**Sara Rubab**

**Fall-2020-MS-EE 327899 SEECS**

Supervisor

**Dr Ahmad Salman**

**Department of Electrical Engineering**

A thesis submitted in partial fulfillment of the requirements for the degree of Masters of  
Science in Electrical Engineering (MS EE)

In

School of Electrical Engineering & Computer Science (SEECS) ,

National University of Sciences and Technology (NUST),

Islamabad, Pakistan.

(July 2024)

## Approval

It is certified that the contents and form of the thesis entitled "Brain Tumor Detection Using Deep Neural Networks" submitted by Sara Rubab have been found satisfactory for the requirement of the degree

Advisor : Dr. Ahmad Salman

Signature:  \_\_\_\_\_

Date: 13-Jun-2024

Committee Member 1: Dr. Wajid Mumtaz

Signature:  \_\_\_\_\_

12-Jun-2024

Committee Member 2: Dr. Salman Abdul Ghafoor

Signature:  \_\_\_\_\_

Date: 13-Jun-2024

Signature: \_\_\_\_\_

Date: \_\_\_\_\_

## THESIS ACCEPTANCE CERTIFICATE

Certified that final copy of MS/MPhil thesis entitled "Brain Tumor Detection Using Deep Neural Networks" written by Sara Rubab, (Registration No 00000327899), of SEECs has been vetted by the undersigned, found complete in all respects as per NUST Statutes/Regulations, is free of plagiarism, errors and mistakes and is accepted as partial fulfillment for award of MS/M Phil degree. It is further certified that necessary amendments as pointed out by GEC members of the scholar have also been incorporated in the said thesis.

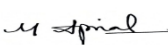
Signature:  \_\_\_\_\_

Name of Advisor: Dr. Ahmad Salman

Date: 13-Jun-2024

HoD/Associate Dean:  \_\_\_\_\_

Date: 13-Jun-2024

Signature (Dean/Principal):  \_\_\_\_\_

Date: 13-Jun-2024

FORM TH-4



**National University of Sciences & Technology**  
**MASTER THESIS WORK**


We hereby recommend that the dissertation prepared under our supervision by: (Student Name & Reg. #) Sara Rubab [00000327899]


Titled: Brain Tumor Detection Using Deep Neural Networks

be accepted in partial fulfillment of the requirements for the award of Master of Science (Electrical Engineering) degree.

**Examination Committee Members**

1. Name: Wajid Mumtaz Signature:   
25-Jul-2024 11:39 AM
2. Name: Salman Abdul Ghafoor Signature:   
25-Jul-2024 11:39 AM

Supervisor's name: Ahmad Salman Signature:   
26-Jul-2024 4:30 PM

  
\_\_\_\_\_  
Salman Abdul Ghafoor  
HoD / Associate Dean

26-July-2024

\_\_\_\_\_  
Date

**COUNTERSIGNED**

\_\_\_\_\_  
29-July-2024  
Date



\_\_\_\_\_  
Muhammad Ajmal Khan  
Principal

THIS FORM IS DIGITALLY SIGNED

Publish Date & Time:

# Dedication

This thesis is dedicated to all the deserving children who do not have access to quality education especially young girls.

## Certificate of Originality

I hereby declare that this submission titled "Brain Tumor Detection Using Deep Neural Networks" is my own work. To the best of my knowledge it contains no materials previously published or written by another person, nor material which to a substantial extent has been accepted for the award of any degree or diploma at NUST SEECs or at any other educational institute, except where due acknowledgement has been made in the thesis. Any contribution made to the research by others, with whom I have worked at NUST SEECs or elsewhere, is explicitly acknowledged in the thesis. I also declare that the intellectual content of this thesis is the product of my own work, except for the assistance from others in the project's design and conception or in style, presentation and linguistics, which has been acknowledged. I also verified the originality of contents through plagiarism software.

Student Name: Sara Rubab

Student Signature: 


**Certificate for Plagiarism**

It is certified that PhD/M.Phil/MS Thesis Titled "Brain Tumor Detection Using Deep Neural Networks" by Sara Rubab has been examined by us. We undertake the follows:

- a. Thesis has significant new work/knowledge as compared already published or are under consideration to be published elsewhere. No sentence, equation, diagram, table, paragraph or section has been copied verbatim from previous work unless it is placed under quotation marks and duly referenced.
- b. The work presented is original and own work of the author (i.e. there is no plagiarism). No ideas, processes, results or words of others have been presented as Author own work.
- c. There is no fabrication of data or results which have been compiled/analyzed.
- d. There is no falsification by manipulating research materials, equipment or processes, or changing or omitting data or results such that the research is not accurately represented in the research record.
- e. The thesis has been checked using TURNITIN (copy of originality report attached) and found within limits as per HEC plagiarism Policy and instructions issued from time to time.

**Name & Signature of Supervisor**

Dr. Ahmad Salman

Signature : 

# Acknowledgments

I would like to thank all my teachers who have taught me starting from school till university for their contribution in building my life and for making me the person I am. I am especially grateful to my supervisor; Dr. Ahmad Salman for his guidance, support, faith and for giving me an opportunity to prove my skills. In the end, I would like to thank SEECs (School of Electrical Engineering and Computer Sciences) for helping me along the way and for giving me an opportunity to attend the best engineering institute in Pakistan.



# Abstract

After cardiovascular disease, cancer is the second major cause of death. Brain tumors really do have the lowest overall survival rate of any type of cancer. Brain tumors are classified according to their morphologically, and location. Appropriate diagnosis of the Tumor type enables the physician to make the best treatment sensible decision and potentially save the patient's life. In the domain of Artificial Intelligence, there is a critical need for a Computer-Aided Diagnosis (CAD) system that can assist physicians and radiologists with diagnosing and classification of cancers. The most powerful and common Machine learning models used for different image analysis tasks like 3D analysis, image retrieval, image classification, and object detection are known as Deep Neural Networks (DNNs). They have achieved a performance level near the human level. Based on the success of DNNs on natural images (e.g., captured images from natural scenes like Imagenet and Cifar10), they have become very popular for tasks such as medical image processing, organ/landmark localization, diagnosis of Cancer, diabetic retinopathy detection, and Covid19 identification. In this study, a novel methodology will be proposed for the early diagnosis and classification of Brain tumors using the different models of DNNs and transfer learning.

# Dedication

This thesis is dedicated to all the deserving children who do not have access to quality education especially young girls.

# Contents

<b>Acknowledgement</b>	<b>vii</b>
<b>Abstract</b>	<b>viii</b>
<b>List of Tables</b>	<b>xiii</b>
<b>List of Figures</b>	<b>xiv</b>
<b>1 Introduction and Motivation</b>	<b>1</b>
1.1 Problem Statement and Contribution . . . . .	3
1.2 Study's Objective . . . . .	3
1.3 Advantages and Potential Applications . . . . .	4
1.3.1 Advantages . . . . .	4
1.3.2 Potential Applications . . . . .	4
<b>2 Literature Review</b>	<b>6</b>
2.1 Background Concepts . . . . .	6
2.2 Machine Learning . . . . .	7
2.2.1 Classification . . . . .	7
2.3 Deep Learning . . . . .	8
2.3.1 Deep Learning in Medical Field . . . . .	8
2.4 Adversarial Attacks . . . . .	9
2.4.1 White Box Attack and White Box Attack . . . . .	9

## CONTENTS

2.4.2	Targeted and Non-targeted Attacks . . . . .	10
2.4.3	Adversarial Attacks Common Type . . . . .	10
2.5	Related Work . . . . .	10
2.5.1	Machine learning for Brain tumor detection . . . . .	10
2.5.2	Deep Neural Network for Brain tumor detection . . . . .	11
<b>3</b>	<b>Deep Learning Models and Transfer Learning</b>	<b>13</b>
3.1	Deep Learning Models . . . . .	13
3.2	Convolutional Neural Network CNN . . . . .	13
3.2.1	Alexnet . . . . .	14
3.2.2	VGGNET . . . . .	15
3.2.3	DenseNet . . . . .	15
3.2.4	Squeezenet . . . . .	16
3.2.5	ResNet . . . . .	16
3.3	Transfer Learning . . . . .	17
<b>4</b>	<b>Proposed Methodology</b>	<b>19</b>
4.1	Proposed method pipeline . . . . .	19
4.2	Dataset . . . . .	20
4.3	Data-Preprocessing . . . . .	21
4.4	Data Augmentation . . . . .	22
4.5	Proposed Model . . . . .	22
4.6	Adversarial attacks . . . . .	23
4.6.1	Fast Gradient Sign Method . . . . .	24
4.6.2	Patch gradient Method . . . . .	25
4.6.3	Basic Iterative Method . . . . .	25
<b>5</b>	<b>Results and Discussion</b>	<b>26</b>
5.1	Results for the Brain Tumor diagnosis . . . . .	26

## CONTENTS

5.2	Performance Metrics	27
5.2.1	Accuracy	27
5.2.2	Recall	27
5.2.3	Specificity	28
5.2.4	Precision	28
5.2.5	F1-score	28
5.3	Parameters Count	28
5.4	Classical Adversarial Attacks on Medical Images	28
5.4.1	Analysis of FGSM based attack on Brain Tumor X-rays	28
5.4.2	Analysis of PGD based attack on Brain Tumor X-rays	29
5.4.3	Analysis of BIM based attack on Brain Tumor X-rays	31
<b>6</b>	<b>Conclusion and Future Work</b>	<b>32</b>
6.1	Future Work	33

# List of Tables

4.1	Hyper-parameters for proposed method . . . . .	23
5.1	Accuracies for different DL models . . . . .	27
5.2	Parameters of different Models . . . . .	29
5.3	Accuracy scores for different models under various epsilon values. . . . .	30
5.4	Accuracy scores for different models under various epsilon values. . . . .	30
5.5	Accuracy scores for different models under various epsilon values. . . . .	31

# List of Figures

3.1	General Structure of CNN	14
3.2	Architecture of Alexnet	15
3.3	Architecture of VGG model	15
3.4	Architecture of DenseNet citeR41	16
3.5	Architecture of SqueezeNet	17
3.6	Architecture of Residual block	18
3.7	Block diagram of Transfer learning	18
4.1	Overall methodology for Classical and Transform domain attacks	20
4.2	Overall methodology for Classical and Transform domain attacks	21
4.3	Block diagram of FSGM	24
4.4	PGD attack	25
5.1	Confusion Matrix	27

## CHAPTER 1

# Introduction and Motivation

Cancer is a major issue worldwide, ranking second only to cardiovascular diseases in causing death and responsible for one-sixth of all global deaths. Brain tumors are one of the most lethal types of cancer, owing to their aggressive nature, diverse traits, and low probability of survival. Brain tumors have various shapes, types, and locations. For instance, Acoustic Neuroma, Meningioma, Pituitary, Glioma, and CNS Lymphoma, among others. In clinical settings, Glioma, Meningioma, and Pituitary tumors make up roughly 45 percent, 15 percent, and 15 percent of all brain tumors. By diagnosing the tumor type, physicians can determine patient survival and choose the best treatment option from surgery, chemotherapy, radiotherapy, and less invasive "wait and see" approaches. Therefore, tumor grading is critical to treatment planning and monitoring [39, 30, 32, 19]. Magnetic Resonance Imaging (MRI) is a non-invasive and painless medical imaging procedure that captures high-quality 2D and 3D images of human body organs, making it one of the most precise methods for detecting and classifying cancer. However, recognizing the type of cancer from MRI images is challenging, specialized, and reliant on the radiologist's experience, making the process prone to errors. Moreover, the tumor's shape may vary with few visible landmarks, making human diagnosis unreliable, which can decrease the patient's survival chances. Conversely, a correct diagnosis can facilitate timely and appropriate treatment, extending the patient's lifespan. Thus, the AI field must design an innovative Computer-Assisted Diagnosis system that can relieve doctors and radiologists of their workload and accurately diagnose and classify tumors.

Recently, the field of Computer-Assisted Diagnosis (CAD) has garnered significant attention for its role in advancing medical imaging and diagnostic radiology. This interest primarily focuses



on improving cancer classification and diagnosis through innovative research and development initiatives. A CAD system usually consists of three primary steps, beginning with the segmentation of lesions from the image. A Computer-Assisted Diagnosis (CAD) system generally involves three main steps. Initially, the system separates the lesions from the image. Subsequently, the model captures the features of the segmented tumors by using innovative statistical or with the help of mathematical parameters learned from a labeled set of MRI images. Lastly, the system uses an appropriate machine learning classifier to anticipate the abnormality class [2, 7]. Before proceeding with classification, many traditional machine-learning techniques necessitate the segmentation of lesions. However, segmentation is a time-consuming and computationally intensive step that can be unreliable and may negatively affect the classification accuracy because of variations in image contrast and intensity normalization. Extracting distinctive parts or features from a raw image, known as feature extraction, is a crucial process in determining the contents of the image. However, this step can be time-consuming and requires prior knowledge of the problem domain. Morphological feature-based classification of tumor types can be misleading since tumors of different types may have a similar appearance. The extracted features are utilized as inputs for machine learning classifiers that classify the image into a specific class based on these features[8]. Unlike traditional machine learning, deep learning does not depend on manually created features. Studies have shown [12, 6] that deep learning has effectively reduced the disparity between human and computer vision when it comes to pattern recognition and can achieve better classification results compared to traditional machine learning methods. The advancements in medical deep learning have made it clear that many state-of-the-art systems can be vulnerable to exploitation by adversarial examples [4, 16]. The risk of adversarial attacks is higher in automated diagnostic processes, as they can stem from various sources such as rare disease image sharing. These attacks involve crafting inputs that deliberately mislead machine learning models into producing inaccurate diagnostic results, known as adversarial examples.

Adversarial attacks on deep learning models often involve altering the classification output of a sample, leading to a decrease in the prediction confidence of the target model. Deep learning models designed for disease diagnosis are more susceptible to these attacks as radiology images are typically pre-defined and easily accessible, making it easier for attackers to manipulate them. This vulnerability is higher in disease diagnosis models compared to other computer vision applications, as stated in source [38]. The risk of successful adversarial attacks on deep-learning

models for rare diseases like COVID-19, Brain tumor, skin cancer and many more diseases is higher due to the limited architectural diversity and data sharing among institutions to generate big data repositories. These models are often publicly available, making it crucial to conduct extensive research to understand potential attacks and develop robust training methods. Adversarial attacks on diagnostic deep-learning algorithms have become a significant concern in both physical and virtual settings, attracting attention and emphasizing the need for research into this area as mentioned in sources [31, 37, 15, 27]. Cutting-edge techniques like FGSM attacks are employed to examine the restrictions of current deep learning methods. These attacks involve optimizing the generation of slight alterations to trick a target model, making it vulnerable to attacks.

## 1.1 Problem Statement and Contribution

Cancer ranks as the second leading cause of death globally, following cardiovascular disease. Among cancers, brain tumors exhibit notably poor survival rates. These tumors are categorized based on their morphology and location. Accurate tumor classification is crucial for physicians to determine the most effective treatment options, potentially saving patients' lives. In Artificial Intelligence (AI), there is a pressing demand for Computer-Aided Diagnosis (CAD) systems to aid physicians and radiologists in cancer diagnosis and classification.

Deep Neural Networks (DNNs) are prominent in AI for various image analysis tasks such as 3D analysis, image retrieval, classification, and object detection. They have demonstrated performance levels approaching human capabilities, particularly in natural image datasets like Imagenet and Cifar10. This success has spurred their popularity in medical applications, including medical image processing, organ localization, cancer diagnosis, diabetic retinopathy detection, and Covid-19 identification. This study proposes a novel methodology using DNNs and transfer learning for early diagnosis and classification of brain tumors.

## 1.2 Study's Objective

The key objective of this research are:

- Diagnosing Brain Tumor via computer aided diagnosis
- Assisting radiologist and doctors in correct diagnosis and subsequent treatment of differ-

ent types of tumors.

- Robustness of deep learning models under normal conditions as well as adversarial attacks.
- Analyzing Brain tumor disease through the lens of deep learning

## **1.3 Advantages and Potential Applications**

### **1.3.1 Advantages**

The following are some of the dissertation's benefits:

- Computer aided diagnosis (CAD) for Brain Tumor
- Early detection of Brain tumor
- Time reduction compared to conventional approach
- Helping physicians at far flung area
- Adversarial attacks on deep learning models
- Transform domain attacks on deep learning models
- Understanding adversarial attacks on Medical Images
- Robustness of Deep learning models towards adversarial attacks

### **1.3.2 Potential Applications**

- Computer-aided diagnosis (CAD)
- Embedded devices
- Artificial intelligence and Deep learning
- Image processing
- Disease diagnosis
- X-rays images

## CHAPTER 1: INTRODUCTION AND MOTIVATION

The remaining sections of this dissertation are as follows: The literature overview is in Chapter 2, followed by deep learning models and transfer learning in Chapter 3, and the methodology of the proposed work is discussed in Chapter 4. In chapters 5 and 6, the results as well as future work are discussed.

## CHAPTER 2

# Literature Review

This chapter looks at Artificial Intelligence based medical diagnostic tools and how they're used. Many countries are still suffering from brain tumors, and while some countries have reported a decline in cases, brain tumors have not been completely eradicated. Various techniques of dealing with brain tumors have been documented in the literature. Researchers are working hard to establish an effective way to diagnose patients and are also frantically trying to find a cure for the disease

### 2.1 Background Concepts

Advances in AI-based methodologies have led in an increase in demand for automatic applications in the medical industry for disease diagnosis over the last few decades [16]. These artificial intelligence (AI) applications aid in enhancing diagnosis accuracy. Many such applications have been developed and are primarily employed in clinics and hospitals in the United States. Artificial intelligence (AI) applications are widely deployed in image processing for object detection, categorization and segmentation. Several of these applications do high-level functions, such as disease prediction. Researchers are primarily focused on the creation of automated systems for tasks that are time demanding for health specialists to complete analysis, hence assisting health professionals in providing better patient care. These AI-based solutions aid pathologists and radiologists tremendously.

## 2.2 Machine Learning

With the increasing amount of data, a tool to analyse it and generate information from it is needed, and Machine Learning can help. Machine learning is a branch of artificial intelligence that assists in the development of autonomous systems in which computers learn about the task at hand without being explicitly coded. Machine Learning may be used in a variety of vocations.

For machine learning algorithms to work, handcrafted features are necessary. Feature selection is a critical stage in Machine Learning, when important characteristics are chosen to allow the model to train and converge

smoothly. In medical diagnostics, a number of machine learning applications have been used. In the pharmaceutical industry, many Machine learning applications have recently been developed, with the system being able to identify patients who are more likely to benefit from the therapies.

The three basic areas of machine learning applications are:

- Supervised Learning
- Unsupervised Learning
- Reinforcement Learning
- The model is given labels and then asked to determine the link between input and output using supervised learning. A supervised model's output can forecast numerous categories where the model is aiming to categorize, and this is referred to as 'Classification' whereas a scalar model's output is referred to as 'Regression'.
- In Unsupervised learning, the labels are not given to the model during training.
- In Reinforcement learning the model learns based on the experiences.

The three essential steps in ML applications are datasets, features, and models. The dataset refers to the massive amount of data that must be mined for understanding. The model is a representation of the phenomenon that a machine learning application has identified, and the features are a subset of this data that aid in the learning process.

### 2.2.1 Classification

If the data is divided into multiple classes or categories, the model will learn the link between the inputs and outputs and attempt to group similar data into a single category. The goal of the

method is to draw distinctions between different kinds of classes. If the supplied data falls under the decision boundary, it will be assigned to that class.

## **2.3 Deep Learning**

Deep Learning is a popular subset of Machine Learning because it does not require feature selection and instead learns features directly from the data, allowing it to handle more complicated tasks. They're popular because they're simple to use and give more accurate results.

Neural networks are used in deep learning. It employs an artificial neural network that attempts to replicate the human brain. Different layers are combined to form a neural network. Neural network is made up of three layers: an input layer, a hidden layer, and an output layer. These levels are made up of 'nodes.' The data is sent into the input layer, the hidden layer does some calculations, and the output layer outputs the required results. The weights and biases of these hidden layers are changed in order to minimize the loss function and allow the model to converge. The depth of a neural network is increased by adding numerous hidden layers, which is why the term "deep" is employed.

Deep learning models work well when there is a lot of data. Depending on the activities they're utilized for. Neural networks come in a number of shapes and sizes. When working with photographs, Convolutional Neural Networks (CNN) are used. You'll want to use a Recurrent Neural Network if you're working on things like Natural Language Processing (RNN). Robotics, video synthesis, facial recognition, and diagnosis of disease are all examples of deep learning applications.

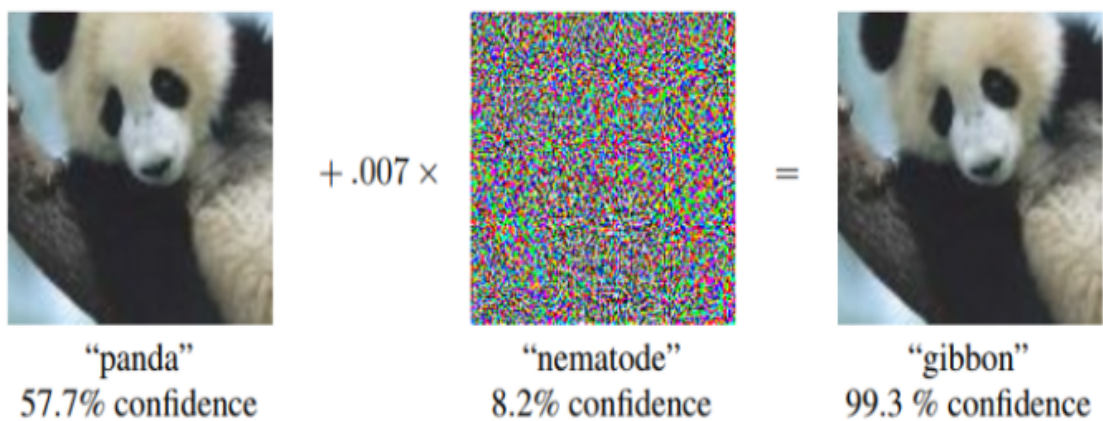
### **2.3.1 Deep Learning in Medical Field**

Because of deep learning, many applications for medical diagnostics have been developed. Deep learning is assisting health professionals in identifying more precise and efficient ways to treat patients. These models aid in medication discovery by examining the patient's medical history, allowing for improved therapy. It can also be used to forecast whether or not a patient's medical insurance claim will be fraudulent. The diseased photos were digitized with full slide scanners and then employed in the deep learning model. This has aided pathologists in analyzing complex

tasks such as cancer detection. These models also make heavy use of radiological scans such as CT and MRI images, which assists radiologists.

## 2.4 Adversarial Attacks

Adversarial attack entails skillfully altering an original image in a way that the modifications remain imperceptible to human vision. The resulting modified image, called an adversarial image, is incorrectly classified by a classifier, while the original image is classified correctly. These attacks can have serious real-world implications; for example, altering a traffic sign could confuse autonomous vehicles, potentially leading to accidents. Another concern is the possibility of illicit content being subtly altered to avoid detection by content moderation algorithms on major websites or by law enforcement web crawlers. The extent of alteration is typically measured using the  $l_1$  norm, which quantifies the maximum absolute shift in a single pixel[29].



**Figure 2.1:** Adversarial Example [38]

### 2.4.1 White Box Attack and White Box Attack

In white-box attacks, the attacker has access to the model's parameters, allowing them to generate adversarial images with precise knowledge of how the model functions. Conversely, in black-box attacks, the attacker lacks access to these parameters. Instead, they create adversarial images using either a different model or no model at all, relying on the expectation that these images will transfer to the target model successfully[26].



## 2.4.2 Targeted and Non-targeted Attacks

Non-targeted attacks aim to induce misclassification of the adversarial image by the model, while targeted attacks aim to manipulate the model into classifying the image as a specific target class different from its true class.

## 2.4.3 Adversarial Attacks Common Type

Gradient-based methods are frequently employed in adversarial attacks. In these methods, attackers adjust the image according to the gradient of the loss function with respect to the input image. There are two main approaches to conducting these attacks: one-shot attacks involve a single adjustment in the gradient direction, while iterative attacks involve multiple successive adjustments instead of a single step[17].

The following are the common types of attack

- Fast gradient sign method
- Targeted Fast gradient method

## 2.5 Related Work

### 2.5.1 Machine learning for Brain tumor detection

Over time, there have been numerous attempts to create an automated system capable of classifying early diagnosis of brain tumors from magnetic resonance image. Several researchers have utilized conventional machine learning techniques, which involve several steps such as pre-processing the images, extracting features, reducing the size of features through feature selection, and applying a classification and detection algorithm to produce the final result. Different methods were used for feature extraction, including Discrete Wavelet and Discrete cosine transform [36]. The techniques used in this context include the Gray Level Co-occurrence Matrix (GLCM) [36], Histogram of Oriented Gradients (HOG) [14], Genetic Algorithm [13], and Zernike Moments. The main difficulty in earlier machine learning research has been the laborious, time-intensive, and error-prone nature of manual feature extraction, which also demands

prior domain expertise. In the past, various approaches have been used to develop an automated system for brain tumor classification using MRI images.

### **2.5.2 Deep Neural Network for Brain tumor detection**

Earlier methods involved segmenting the tumors based on regions before feature extraction as well as classification. Additionally, feature selection necessitated the further reduction of selected features, and there was no single method of extraction of feature that could be universally applied. However, with the advent of deep learning [20], a sub field of machine learning, manual feature engineering is no longer required. Nevertheless, to obtain better outcomes, it should be applied appropriately to preprocessed data, with the use of suitable architectures and hyperparameters. Convolutional Neural Networks (CNNs) are a noteworthy instance of deep learning methods, owing to their advanced image processing capabilities and faster computation rate, they are widely used in brain tumor research. Numerous studies have attempted to leverage CNNs for the diagnosis of brain cancer, as documented in the literature [23, 35]. These efforts seek to determine the optimal model and network architecture for enhancing the automatic classification of brain cancer. Earlier studies on brain tumor classification using deep learning methods utilized smaller datasets compared to the dataset used in this study. This is because medical datasets are relatively rare and difficult to collect. A deep learning model based on CNNs efficiently addressed the problem of brain tumor classification [25]. One of the benefits of a CNN-based classifier system is its ability to provide a fully automated classifier without necessitating manually segmented tumor regions. Pashaei et al. [24] conducted a research in which they developed a CNN structure that extracted brain MRI features. This CNN design had five learnable layers, with all layer filters measuring 3x3. The CNN model achieved an accuracy rate of 81 for classification. However, the accuracy was enhanced by combining CNN features with an extreme learning machines (ELM) classifier model. The classifier's discrimination ability was shown to be limited by the low recall measures for meningioma. To overcome this issue, Afshar et al. [21] utilized a modified CNN structure called CapsNet, which considered the spatial relationship between the tumor and its neighboring tissues.

However, the resulting enhancement in performance was only minor. In studies conducted by other researchers [33], Convolutional Neural Networks were utilized to classify brain tumors,

resulting in better accuracy. On the other hand, authors of [22] introduced a Capsule Networks architecture. This approach resulted in accuracy rates of 90 and 86, respectively. A study by A. Pashaei and colleagues [34] employed a combination of a Convolutional Neural Network and other machine learning techniques, such as KELM, to achieve an accuracy rate of 93.8.

# Deep Learning Models and Transfer Learning

This section provides a detailed explanation of the architecture of deep learning models, including their key components and structure. Additionally, it delves into the concept of transfer learning, discussing how pre-trained models on large datasets can be adapted to specific tasks, improving performance and efficiency by leveraging previously learned features and knowledge.

## 3.1 Deep Learning Models

Deep learning is an emerging field which consists of following basic models.

- Supervised models
- Unsupervised models

In Supervised models, the labels are given to the model during training stage. The most common examples of supervised models are Multilayer Perceptron and Convolutional Neural Network (CNN) while in Unsupervised models, the labels are not given to the models during training stage. The input is fed into the model and the model tries to make prediction without the labels.

## 3.2 Convolutional Neural Network CNN

When dealing with pictures or Computer Vision tasks, a convolutional neural network (CNN) model is used. Numerous CNN applications have been created to assist with object detection,

classification, and segmentation for applications such as human face detection, vehicle identification, and so on. CNN is often utilised in the medical field to aid in disease diagnosis

There are mainly three building blocks in the architecture of CNN

- Convolution Layer: This layer is used for learning the features of an input image.
- Max-Pooling layer: It is also known as subsampling layer. It down samples the image in order to reduce image's dimensionality leading to lesser computational complexity.
- Fully Connected Layer: It is known to instill classification capacities in the network.

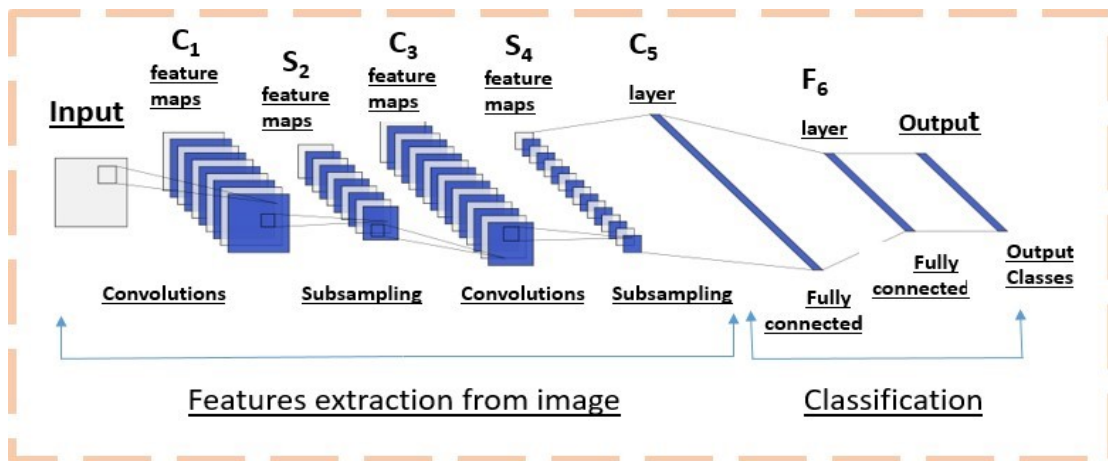


Figure 3.1: General Structure of CNN

### 3.2.1 Alexnet

AlexNet was suggested in 2012 and was the winner of the 2012 ImageNet classification contest. It was the first time that a deep neural network was used to classify images in an imagenet. AlexNet has eight layers in total, excluding the pooling levels. Maxpooling is the pooling method employed. Five convolutional layers and three fully linked layers comprise this image. It employs a wide receptive field (11 x 11) and a small receptive field (5 x 5) in the initial layers and a smaller receptive field (3x3) in the subsequent layers. Each convolutional layer is followed by a ReLU activation function, and the final layer classifies 1,000 classes using the softmax function [18]. The architecture of AlexNet is shown

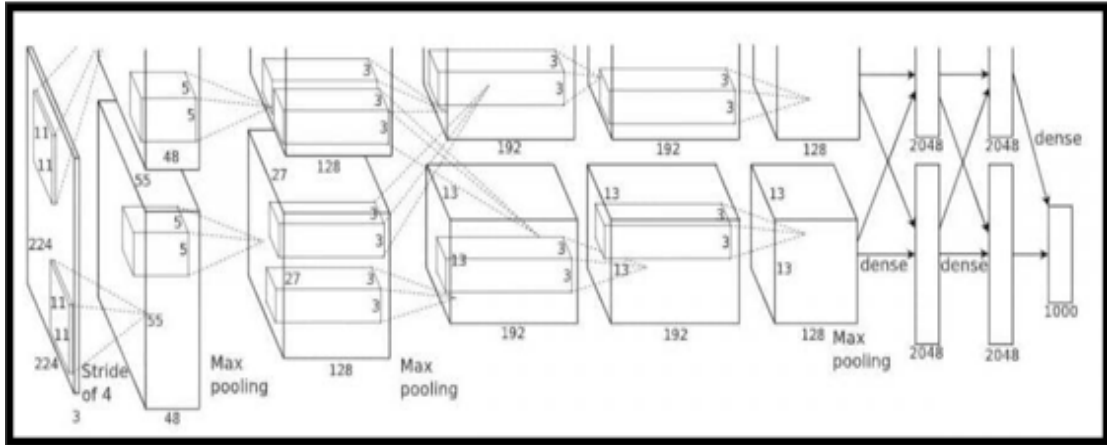


Figure 3.2: Architecture of Alexnet

### 3.2.2 VGGNET

The Oxford Visual Geometry Group developed the VGG architecture, which is the state-of-the-art model in 2014. VGG was an evolution of the AlexNet architecture. VGG has fewer parameters than AlexNet due to the usage of a set of filters with 12 tiny receptive fields of size (3 x 3). On imagenet data, it achieved a top 5 test accuracy of 97.2 percent for the Classification challenge. Vgg has two different variants. One is Vgg-16 which is 16 layer dense model while the other one is Vgg-19 which is 19 layers dense network[28].

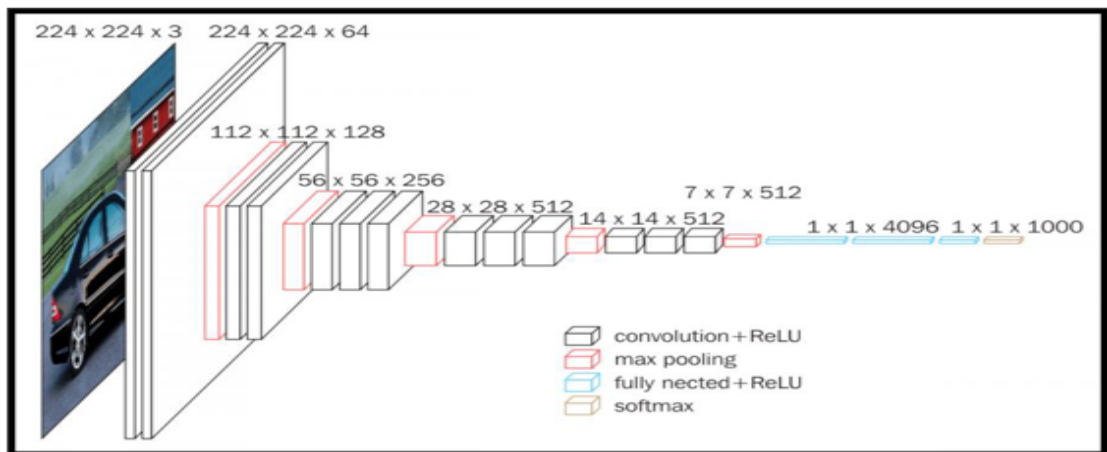


Figure 3.3: Architecture of VGG model

### 3.2.3 DenseNet

Dense Convolutional Networks (DenseNet) require fewer parameters than typical CNNs since they never learn redundant mapping features. DenseNet’s layers are relatively thin, consisting of

12 filters, which result in a smaller collection of new feature maps. DenseNet is available in four flavours: DenseNet121, DenseNet169, DenseNet201, and DenseNet264. The computational cost is minimised since each dense block is directly connected to the input image and loss function gradient [1].

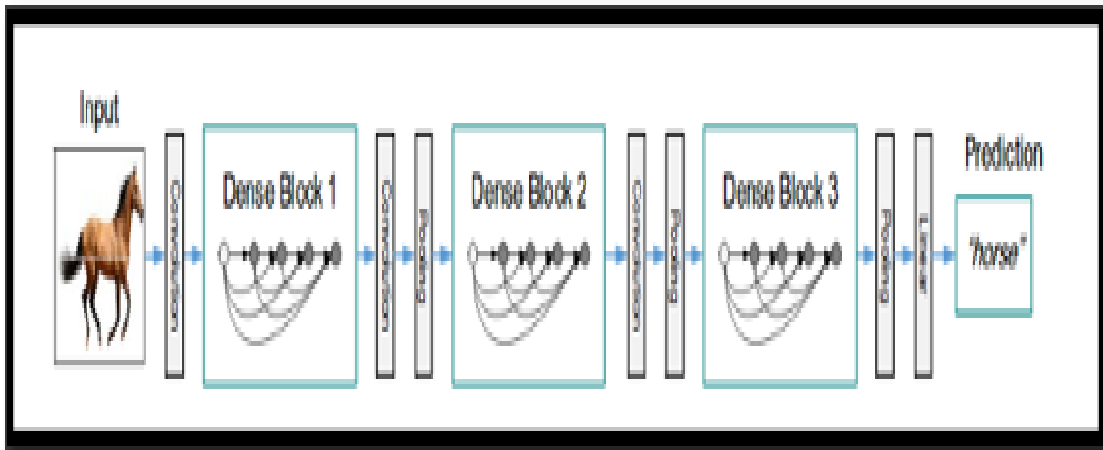


Figure 3.4: Architecture of DenseNet citeR41

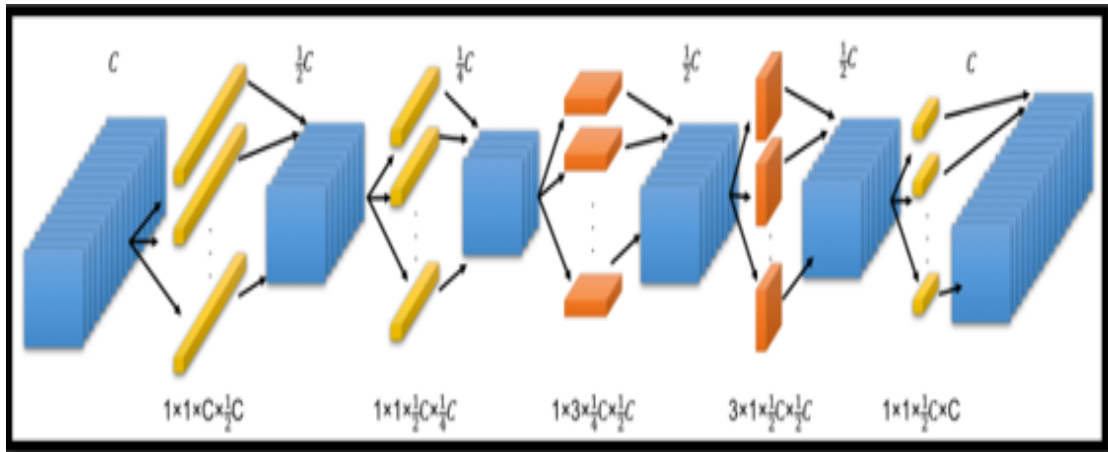
### 3.2.4 Squeezenet

Squeezenet is another sort of CNN that is learned using an imagenet dataset. AlexNet is more sophisticated by a factor of fifty. The network is formed by an extruded and extended fireproof module. The extruded layer contains only a single 11-filter, which is more suitable for an enlarged layer than a mixture of 11 and 33-filters. CNN classifiers are regarded as a significant evolutionary step in the development of Inception models. They are inherently complicated. There are numerous versions, with each one improving on the preceding one iteratively [9].

### 3.2.5 ResNet

The number of layers in deep learning models enhances the network's depth. However, when the network's depth increases, it runs into problems, culminating in a network that is extremely bad and inaccurate. The following are the most common challenges that a deep network encounter:

- Exploding/Vanishing Gradient: As the number of layers in the network rises and the weights are changed, gradients become unstable. As a result of continuous multiplication, the gradient value may climb to an infinitely large value or shrink to an eternally small value, and the weights are not updated.



**Figure 3.5:** Architecture of SqueezeNet

- **Network Degradation:** Another issue is that as you get deeper into the model, its performance begins to deteriorate. The network's accuracy suffers as a result of low performance.

Microsoft created a network called Residual Neural Network (ResNet) to overcome the challenges outlined[10].

The ResNet architecture's basic concept is to use skip connections after numerous levels. After a few more levels, the output from the preceding layers can be applied as is. As a result of the increase in depth, the exploding /vanishing gradient problem is avoided, as well as network performance decrease. These leftover blocks are simply put together to increase the depth of the networks. The input and output dimensions are the same in an identity skip connection. Resnet has been released in many versions, such as resnet50, resnet101, and resnet151. The only distinction between them is the number of layers.

### 3.3 Transfer Learning

When faced with a huge dataset, it is well established that deep transfer learning CNNs outperform smaller networks. As a result, transfer learning is frequently used when smaller datasets are available. The following figure can assist in comprehending the concept of transfer learning. Without sacrificing efficiency, a model trained from a bigger dataset (ImageNet) can be put to use for smaller datasets. Transfer learning has recently been used to for many image recognition tasks as well as for medical images for disease diagnosis[8].



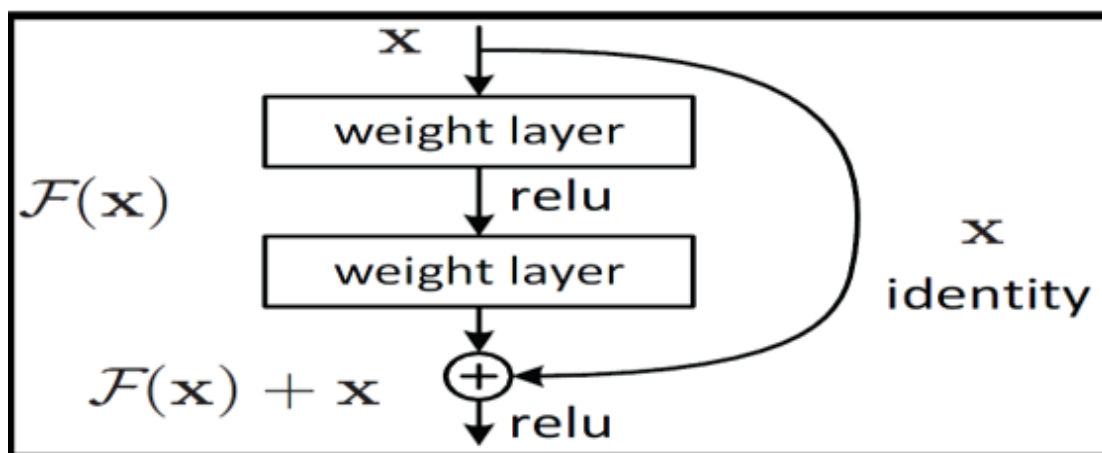


Figure 3.6: Architecture of Residual block

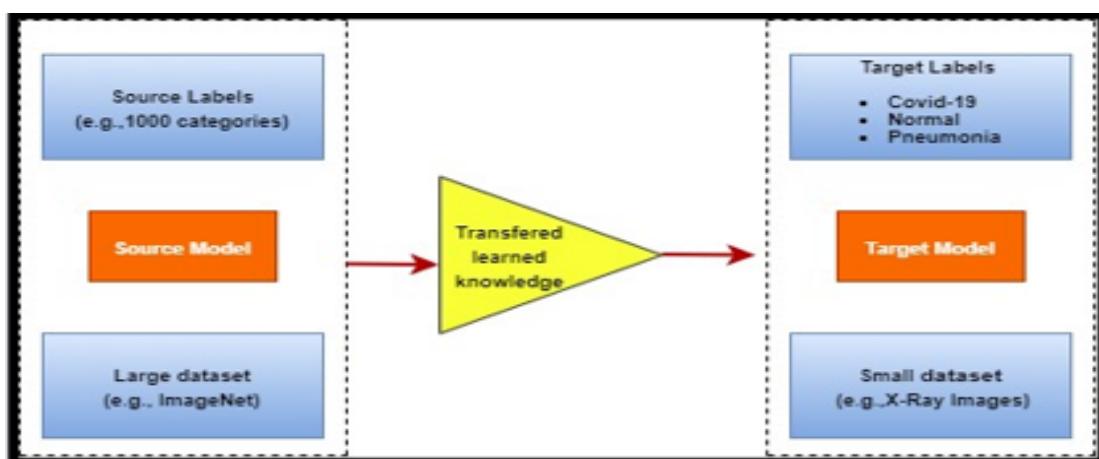


Figure 3.7: Block diagram of Transfer learning

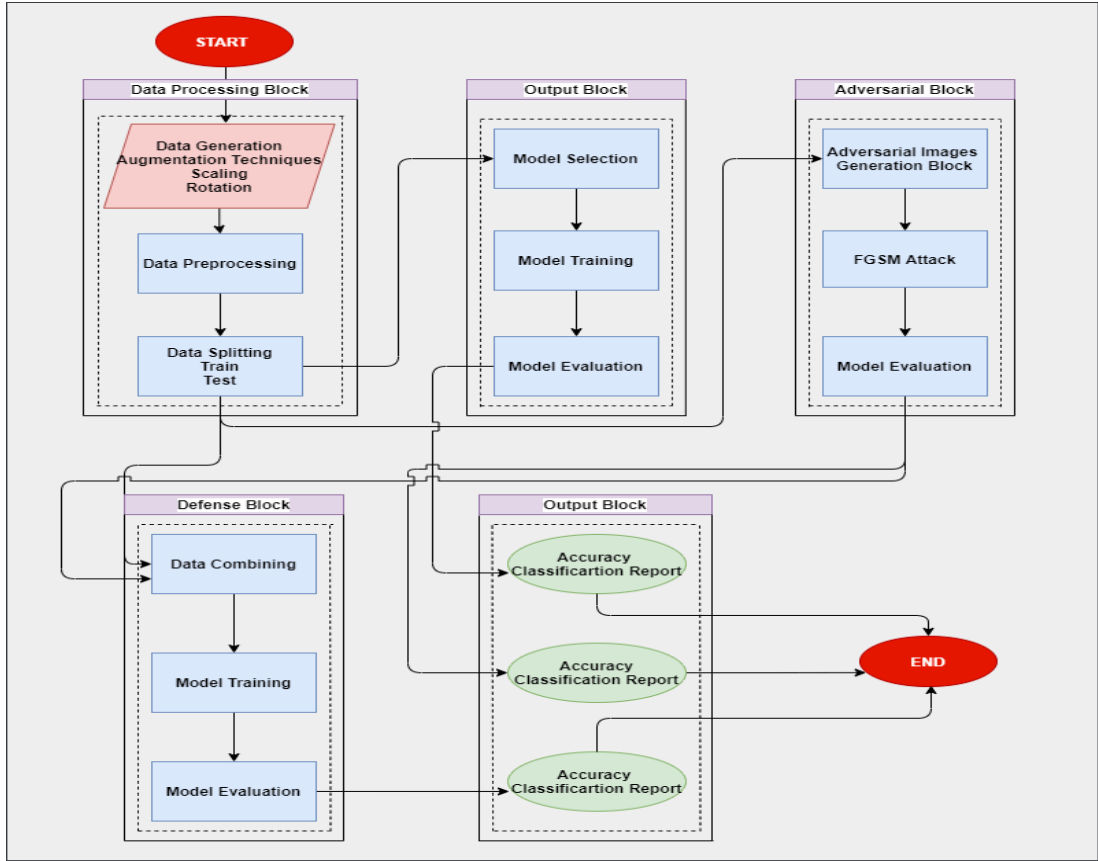
# Proposed Methodology

The following section outlines the proposed model's design and implementation.

## 4.1 Proposed method pipeline

Our methodology comprises several stages that together form the pipeline for constructing our deep learning model. It begins with acquiring raw data and concludes with generating the output. Each stage's output serves as input for the subsequent stage. The proposed approach pipeline is depicted in the figure below, with detailed explanations provided in the following sections[3].

- **Acquiring Data:**Data acquisition is essential for defining the task and evaluating the model's performance. Once we acquired the brain tumor dataset, we conducted necessary image preprocessing and conversions to optimize the model's performance.
- **Preprocessing Data:** Due to constraints, medical image datasets are usually smaller than datasets in other domains. Maximizing the data during runtime and enable the model to achieve excellent results and generalize, we used various augmentation techniques while preprocessing the data.
- **Displaying data in a visual format:** To understand the patterns in the data, we created visualizations of the training data during both the pre-processing and augmentation stages.
- **Creating a Model:** A model refers to an algorithm which receives input data  $X$  and generates predictions for  $Y$  outcome.
- **Model training:** This stage entails optimizing model parameters and updating classification weights through iterative training.



**Figure 4.1:** Overall methodology for Classical and Transform domain attacks

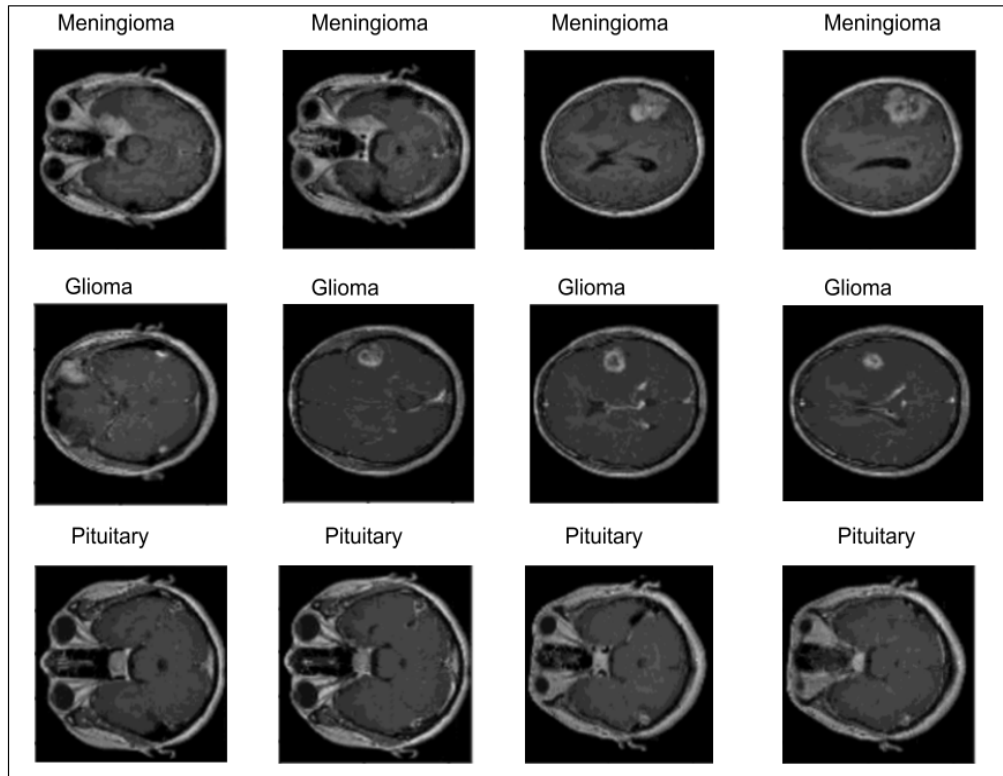
- Evaluating a Model: To evaluate the performance of our model, we used various metrics such as accuracy, precision, recall, f1-score, and balanced accuracy.
- Adversarial attack: In order to analyze the effect of deep learning models on adversarial attacks, we apply different attacks.

## 4.2 Dataset

We will utilize a brain tumor dataset consisting of 3064 T1-weighted contrast-enhanced MRI images, which were made publicly available and collected from Nanfang Hospital in Guangzhou, China, and General Hospital, Tianjin Medical University in China between 2005 and 2010. The dataset encompasses three types of brain tumors: Meningiomas, Gliomas, and Pituitary tumors, which are among the most prevalent brain tumors. Cheng et al. [8] initially processed the dataset to develop a brain tumor classification model. The dataset includes 2D MRI scan images from 233 cancer patients, with each patient's identity anonymized. Figure 5 shows sample images

Category of Tumor	Patients Number	Slices Number
Meningioma Tumor	82	708
Pituitary Tumor	62	930
Glioma Tumr	89	1426

with class labels, and Figure 4 depicts the distribution of images across the different classes.



**Figure 4.2:** Overall methodology for Classical and Transform domain attacks

### 4.3 Data-Preprocessing

Before inputting images into our classifiers, we utilize a variety of preprocessing techniques. For instance, the MRI images in the Figshare dataset are in a ".mat" format defined in Matlab, so we must expand the dimensions of the image to read it. Subsequently, we convert all images into NumPy arrays, which take up less space and are compatible with our model in Python. To ensure our model can train on unordered data, we shuffle the dataset before splitting it into three sections: training, testing, and validation. We allocate approximately 80 of the data to training

and the remaining 20 to validation and testing.

## 4.4 Data Augmentation

To address the issue of inadequate training data, one effective solution is to utilize data augmentation techniques. By applying modifications such as adjusting brightness, scaling, and flipping to the existing data, new images with the same label can be created, effectively expanding the dataset. This approach is particularly useful for deep learning models, which perform better with larger datasets. Data augmentation also acts as a form of regularization on the dataset level, reducing overfitting and improving generalization performance without modifying the model architecture. Additionally, data augmentation can help address the class imbalance by over-sampling the minority class, resulting in more balanced training data. Medical image datasets are often limited in size and difficult to obtain, making data augmentation especially useful in applications such as skin lesion classification as well as disease of liver lesion detection and classification and brain scan analysis and other medical imaging tasks [11, 3].

In our study, we incorporated diverse augmentation methods during the training process to increase the number of images. These techniques included flipping the images horizontally and vertically as well as rotating them.

## 4.5 Proposed Model

In our groundbreaking research, we employed a comprehensive ensemble of eight distinct models, each meticulously crafted to tackle the intricate challenge of brain tumor classification. Among these formidable models are our bespoke Convolutional Neural Network (CNN), renowned architectures like AlexNet, ResNet-34, SqueezeNet, GoogLeNet, DenseNet-121, Inception-V3, and the venerable VGG-16.

Harnessing the power of these models demanded extensive training on our meticulously curated dataset. With 3,216 high-resolution images earmarked for training and a further 805 meticulously selected images for rigorous evaluation, our models were put through their paces.

To overcome the inherent limitations of training data and time constraints, we adopted a strategic transfer learning approach. By leveraging pre-trained models, we expedited the learning

Hyper-Parameter	Optimized Values
Optimizer used	SGD
Activation function used	Relu
Learning rate used	0.0001
Batch size used	32
Epochs used	30

**Table 4.1:** Hyper-parameters for proposed method

process while ensuring adaptability to our specific task. This entailed freezing the upper layers of the pre-trained models and introducing fully connected layers tailored to our dataset. Through this meticulous process, we aimed to extract nuanced features while preserving the invaluable insights encapsulated within the pre-trained weights.

Our training regimen spanned an exhaustive 300 epochs, with the Rectified Linear Unit (ReLU) activation function and Stochastic Gradient Descent (SGD) optimizer at the helm. SGD optimization, renowned for its stochastic nature, facilitated dynamic adjustments to model parameters, optimizing convergence towards our desired outcomes.

For transparency and reproducibility, we have meticulously documented the hyperparameters governing our model's performance, as outlined in table below. Each parameter meticulously calibrated to strike the delicate balance between model complexity and performance, ensuring robustness and generalizability across diverse datasets.

## 4.6 Adversarial attacks

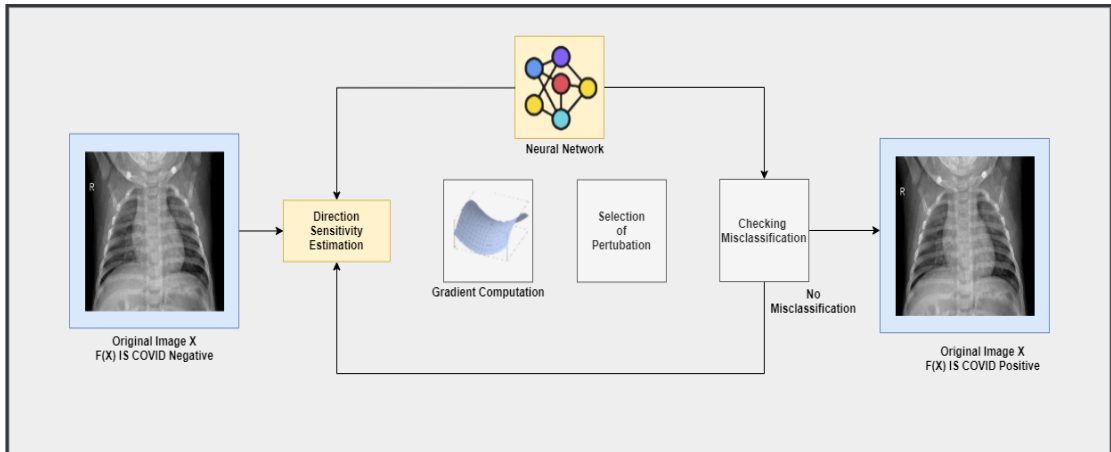
An adversarial attack refers to a subtle modification of an original image that is almost imperceptible to the human eye, resulting in an adversarial image that is misclassified by the classifier. This type of attack, known as adversarial noise, can greatly compromise the robustness of deep neural networks used for various image classification tasks. There are two categories of adversarial attacks: in-distribution (IND) and out-of-distribution (OOD) [31]. Although there have been extensive studies on IND adversarial attacks for a variety of applications, this research illustrates that attacks such as FGSM can effectively impair the performance of dependable DL models [32,33].

### 4.6.1 Fast Gradient Sign Method

In a text, the fast gradient sign method creates an adversarial example using the neural network's gradients. The fast gradient sign process for determining adversarial images was invented by Ian Goodfellow et al. (2014) [5]. According to Equation (1), the gradient sign method generates adversarial examples by using the gradient of the existing model:

$$y' = y + \varepsilon \cdot \text{sign}(\nabla_y J(\theta, y, z)) \quad (4.6.1)$$

The input image is denoted by  $y$ , and its original class is denoted by  $z$ , and the parameter vector of model is denoted by  $\theta$ .  $J(\theta, y, z)$  denotes the loss function that was used in training of the network. To begin, with the loss function gradient the is calculated through the input pixels. The operator  $\nabla$  is a mathematical technique for computing the function's derivatives with respect to the model's various parameters. As a result,  $\nabla_x J(\theta, y, z)$  is now the gradient vector through which the sign is obtained. The gradient can have either a positive or negative sign, based on the loss function. The positive sign indicates that loss increases by increasing the pixel intensity, i.e., the model's error, while the negative sign indicates that by decreasing the pixel intensity results in increasing the loss. This vulnerabilities occur when the model treats the relationship between the intensity of an input pixel and the class score linearly. Figure 3 illustrates the procedure. The sign ( $xJ(,y,z)$ ) represents the product of a very slight epsilon value and the signed values



**Figure 4.3:** Block diagram of FSGM

calculated through the gradient vector. The output of the multiplication is then added to that original image  $Y$  to generate the adversarial images  $Y$ .

$$y' = y + \eta \quad (4.6.2)$$

where  $\eta$  denotes  $\text{sign}(yJ(\theta, y, z))$ .

Thus, differing the value of epsilon, which is typically between 0 and 1, results in a variety of adversarial examples. The majority of these examples are imperceptible to the naked eye [34].

#### 4.6.2 Patch gradient Method

It is also a white box attack meaning that all the model gradients can be accessed by the attacker. In other words, the attacker is equipped with a copy of weights of the neural network. White box attacks are way more powerful than other attacks as they give access to the attacks to craft the attack for fooling the model in such a way that the perturbations are discarded which can be evident in case of transfer attacks. PGD is considered as the strongest white box attack as it minimizes the effort and time for an attacker into finding the best attack.



Figure 4.4: PGD attack

#### 4.6.3 Basic Iterative Method

This attack is an advanced version of FGSM. In order to create this attack, step size of FGSM is made smaller and it is replaced multiple times. The result obtained after each iteration is clipped so that perturbation stays within the original image's neighborhood. Resultantly, basic iterative method becomes much stronger step having small perturbations.



# Results and Discussion

This chapter delves into the software utilized for implementation purposes, along with a detailed examination of the recommended approach and its outcomes. Furthermore, it scrutinizes the performance of the proposed work, specifically focusing on early diagnosis of brain tumors utilizing Deep Neural Networks (DNN), as well as classical and transform domain adversarial attacks on medical images.

Pytorch enables the end user to readily access a variety of deep learning models and pre-trained weights. It enables significant time savings by avoiding the need to develop and train deep learning models from scratch. The input size is reduced to different dimensions to make in compatible for different deep learning models.

## 5.1 Results for the Brain Tumor diagnosis

In this section, we conducted a comparative analysis of various pre-trained models alongside our custom-designed model. The training phase utilized 3216 data samples, with an additional 804 samples reserved for model evaluation. Results indicate that fine-tuned versions of AlexNet, Inception V3, and SqueezeNet achieved the highest accuracy of 98% on test images. Our proposed model also performed well, achieving an accuracy of 97.02%, notable for its efficiency with fewer computational demands and parameters compared to other models. The performance of these models is summarized in the table below.

Parameter	Values
BrainNet (self-designed)	97.02
ResNet-18	94
AlexNet	92
VGG-19	93
Inception v3	98

**Table 5.1:** Accuracies for different DL models

## 5.2 Performance Metrics

A simple confusion matrix is shown below. Five different criteria are explained below that

		Predicted Class	
		Yes	No
Actual Class	Yes	True Positive	False Negative
	No	False Positive	True Negative

**Figure 5.1:** Confusion Matrix

encompasses the techniques of deep transfer learning:

### 5.2.1 Accuracy

It is the ratio of True Positive and True Negative observations to the total number of observations.

$$(t_p + t_n) / (t_p + t_n + f_p + f_n) \quad (5.2.1)$$

where  $t_n$  and  $t_p$  is for true negative and true positive respectively while  $f_p$  represents false positive and false negative is shown by  $f_n$ .

### 5.2.2 Recall

Recall is calculated as

$$t_p / (t_p + f_n) \quad (5.2.2)$$

### 5.2.3 Specificity

In general specificity is calculated by

$$t_n / (t_n + f_p) \quad (5.2.3)$$

### 5.2.4 Precision

It compares the True Positive to all positive cases in the Predicted class.

$$t_p / (t_p + f_p) \quad (5.2.4)$$

### 5.2.5 F1-score

It is determined by the use of Recall and Precision.

$$2x[(precision \times recall) / (precision + recall)] \quad (5.2.5)$$

The F1-Score is directly proportional to the classifier's performance. A high F1-score indicates that the classifier is doing well.

## 5.3 Parameters Count

The table below summarises the entire number of parameters in our custom-built CNN and various models. VGG16 contains the most parameters, indicating that it is more computationally complex than other models. In comparison to previous pre-trained models, our suggested CNN model has less parameters.

## 5.4 Classical Adversarial Attacks on Medical Images

This section explains the generation of adversarial images in classical domain using the concept of FGSM in order to fool the state of art DNN.

### 5.4.1 Analysis of FGSM based attack on Brain Tumor X-rays

In our study focusing on brain tumor sample categorization, we employed transfer learning-based models to enhance classification accuracy. However, to evaluate the robustness of these

Parameter	Values
ResNet 34	21.282 million
CNN (Self designed)	13 million
AlexNet	61 million
GoogleNet	7 million
VGG 16	138 million
Inception V3	24 million
DenseNet 121	7.2 million
SqueezeNet	7 million

**Table 5.2:** Parameters of different Models

models, we subjected them to the Fast Gradient Sign Method (FGSM) attack. This attack is a commonly used technique to generate adversarial examples by perturbing input data slightly in the direction of the gradient of the loss function with respect to the input.

The primary objective was to investigate the impact of FGSM perturbations on the perceptibility of brain MRI images and to determine whether such perturbations could lead to misclassification by both human and machine radiologists. By analyzing the FGSM results for both binary and multiclass scenarios, we aimed to understand the vulnerability of the models to adversarial attacks and assess the potential risks associated with deploying them in clinical settings.

Our findings shed light on the susceptibility of transfer learning-based models to adversarial attacks, highlighting the need for robust defenses to safeguard against such threats in medical image analysis applications.

From the Table 5.3, it can be seen that by increasing the value of epsilon, the perturbations increase as a result the accuracies of different deep learning models drop abruptly. For high values of epsilon, the perturbations can be easily seen by the naked eye.

#### 5.4.2 Analysis of PGD based attack on Brain Tumor X-rays

Building upon our investigation into the robustness of transfer learning-based models for brain tumor classification, we extended our analysis to include the Projected Gradient Descent (PGD) attack. Unlike the FGSM attack, PGD iteratively applies small perturbations to input data while ensuring that the perturbed samples remain within a specified epsilon-bound neighborhood. This

**Table 5.3:** Accuracy scores for different models under various epsilon values.

Model	Accuracy	Epsilon 0	Epsilon 0.0002	Epsilon 0.0004	Epsilon 0.0006	Epsilon 0.005
Inception-V3	97	97	95	93	91	37
AlexNet	92	92	91	90	89	54
ResNet-18	84	84	21	5	1.6	3.2
Vgg19	87	87	82	75	70	2.2
BrainNet	97.02	97.02	86	80	65	40

approach allows for a more systematic exploration of the model's vulnerability to adversarial manipulation.

Similar to our evaluation with the FGSM attack, our focus with PGD was twofold: first, to assess the degree of perturbation required to mislead the models, and second, to evaluate the perceptibility of the perturbed brain MRI images. By subjecting the models to PGD attacks and analyzing the resulting adversarial examples, we aimed to gain deeper insights into the resilience of the models against adversarial perturbations and the potential impact on their performance in clinical settings.

Through our investigation of both FGSM and PGD attacks, we aim to contribute valuable insights into the robustness and reliability of transfer learning-based models for brain tumor classification. Our findings will inform the development of more resilient models and defense mechanisms to mitigate the risks associated with adversarial attacks in medical image analysis.

From the table below, it can be seen that by increasing the value of epsilon, the perturbations increase as a result the accuracies of different deep learning models drop abruptly. For high values of epsilon, the perturbations can be easily seen by the naked eye.

**Table 5.4:** Accuracy scores for different models under various epsilon values.

Model	Accuracy	Epsilon 0	Epsilon 0.05	Epsilon 0.07	Epsilon 0.105	Epsilon 0.1
Inception-V3	97	97	24	24	24	24
AlexNet	92	92	54	36	32	32
ResNet-18	84	84	33	32	31	31
Vgg19	87	87	85 (0.0001)	26 (0.002)	7 (0.05)	7 0.1
BrainNet	97.02	97.02	81	50	40	25

### 5.4.3 Analysis of BIM based attack on Brain Tumor X-rays

In addition to evaluating the susceptibility of transfer learning-based models to adversarial attacks like the FGSM, our study also explored the implications of the Boundary-Input-Masking (BIM) attack on brain tumor classification. The BIM attack extends the FGSM approach by iteratively applying small perturbations to input data, aiming to maximize the model's misclassification rate while ensuring the perturbations remain imperceptible to human observers.

By subjecting our models to the BIM attack and analyzing the resulting perturbed brain MRI images, we sought to ascertain the effectiveness of this attack strategy in circumventing classification models trained on medical imaging data. Our investigation focused on quantifying the extent of perturbation required to induce misclassification and assessing the potential impact of such attacks on the reliability of automated brain tumor diagnosis systems.

Through comprehensive analysis of the BIM attack results, we aimed to enhance our understanding of the vulnerabilities inherent in transfer learning-based models deployed in medical image analysis tasks. These insights are crucial for developing robust defense mechanisms to mitigate the risks posed by adversarial attacks in clinical settings.

From the table below, it can be seen that by increasing the value of epsilon, the perturbations increase as a result the accuracies of different deep learning models drop abruptly. For high values of epsilon, the perturbations can be easily seen by the naked eye.

**Table 5.5:** Accuracy scores for different models under various epsilon values.

Model	Accuracy	Epsilon 0	Epsilon 0.05	Epsilon 0.01	Epsilon 0.105	Epsilon 0.2
Inception-V3	97	97	47	47	47	47
AlexNet	92	92	64	64	64	64
ResNet-18	84	84	30	30	30	30
Vgg19	87	87	6	6	6	6
BrainNet	97.02	97.02	90	81	70	50

# Conclusion and Future Work

The concluding remarks of our study on early detection methods for brain tumors and the evaluation of adversarial attacks on medical images using classical techniques are elaborated upon in this chapter. We delve into the effectiveness and limitations of our proposed approach in accurately identifying brain tumors from MRI scans, considering both the potential benefits and challenges posed by adversarial attacks.

Through comprehensive analysis of our experimental results, we shed light on the strengths and weaknesses of our detection framework, highlighting its performance in differentiating between tumor and non-tumor regions in brain images. Furthermore, we discuss the impact of adversarial attacks on the robustness of our classification models, emphasizing the need for robust defense mechanisms to counter such attacks.

In addition to summarizing our findings, we outline potential avenues for future research in this domain. These include exploring novel techniques for enhancing the resilience of brain tumor detection models against adversarial attacks, investigating the integration of multi-modal imaging data for improved diagnostic accuracy, and exploring the feasibility of deploying our approach in clinical settings.

By addressing these research directions, we aim to contribute to the ongoing efforts in advancing early detection methods for brain tumors while addressing the emerging challenges posed by adversarial attacks on medical image analysis systems. Through collaboration and continued exploration of innovative methodologies, we strive to improve the effectiveness and reliability of brain tumor diagnosis in clinical practice.

## **6.1 Future Work**

It is imperative to detect brain tumor patients early to mitigate the progression of the disease. In this research, we have proposed a method called deep transfer learning, which leverages MRI images to identify patients with brain tumors and predict whether the tumor is present or not on a diagnostic basis. Our findings suggest that this technique can aid healthcare professionals in decision-making processes in various ways. Moreover, by augmenting the brain tumor dataset with a larger repository containing diverse chest disorders data, we can develop a more robust and practical model. Additionally, subtle modifications to the dataset through adversarial attacks in both classical and transformed domains can potentially deceive deep learning algorithm.



# Bibliography

- [1] G. Simon. *Principles of Chest X-ray Diagnosis*. Butterworths, Oxford, MA, USA, 1971.
- [2] Alex Krizhevsky, Ilya Sutskever, and Geoffrey E Hinton. “Imagenet classification with deep convolutional neural networks”. In: *Advances in neural information processing systems*. 2012, pp. 1097–1105.
- [3] Alex Krizhevsky, Ilya Sutskever, and Geoffrey E. Hinton. “Imagenet classification with deep convolutional neural networks”. In: *Advances in neural information processing systems*. 2012, pp. 1097–1105.
- [4] Ian J. Goodfellow, Jonathon Shlens, and Christian Szegedy. “Explaining and harnessing adversarial examples”. In: *arXiv preprint arXiv:1412.6572* (2014).
- [5] D. P. Kingma and J. Ba. “Adam: A method for stochastic optimization”. In: *arXiv preprint arXiv:1412.6980* (2014).
- [6] Christian Szegedy et al. “Intriguing properties of neural networks”. In: *arXiv preprint arXiv:1312.6199* (2014).
- [7] M. M. Najafabadi et al. “Deep learning applications and challenges in big data analytics”. In: *Journal of Big Data* 2.1 (2015), p. 1.
- [8] A. Kurakin, I. Goodfellow, and S. Bengio. “Adversarial Examples in the Physical World”. In: *arXiv preprint arXiv:1607.02533* (2016).
- [9] N. Papernot et al. “Distillation as a defense to adversarial perturbations against deep neural networks”. In: *Proceedings of the 2016 IEEE Symposium on Security and Privacy (SP)*. IEEE. 2016, pp. 582–597.
- [10] K. Grosse et al. “Adversarial Examples for Malware Detection”. In: *Proceedings of the 22nd European Symposium on Research in Computer Security*. Springer. 2017, pp. 62–79.

## BIBLIOGRAPHY

- [11] M. Melis et al. “Is deep learning safe for robot vision? Adversarial examples against the iCub humanoid”. In: *Proceedings of the IEEE International Conference on Computer Vision Workshops*. 2017, pp. 751–759.
- [12] D. Shen, G. Wu, and H.-I. Suk. “Deep learning in medical image analysis”. In: *Annual review of biomedical engineering* 19 (2017), pp. 221–248.
- [13] Kenji Suzuki. “Overview of deep learning in medical imaging”. In: *Radiological Physics and Technology* 10.3 (2017), pp. 257–273.
- [14] David Ardila et al. “End-to-end lung cancer screening with three-dimensional deep learning on low-dose chest computed tomography”. In: *Nature Medicine* 25 (2019), pp. 954–961.
- [15] Jahanzaib Latif et al. “Medical imaging using machine learning and deep learning algorithms: a review”. In: *2019 2nd International Conference on Computing, Mathematics and Engineering Technologies (iCoMET)*. IEEE. 2019, pp. 1–7.
- [16] Jiawei Su, Danilo V. Vargas, and Kouichi Sakurai. “One pixel attack for fooling deep neural networks”. In: *IEEE Transactions on Evolutionary Computation* (2019).
- [17] Tao Ai et al. “Correlation of Chest CT and RT-PCR Testing for Coronavirus Disease 2019 (COVID-19) in China: A Report of 1014 Cases”. In: *Radiology* 296.2 (2020). PMID: 32101510; PMCID: PMC7233399. doi: 10.1148/radiol.2020200642, E32–E40.
- [18] Luca Brunese et al. “Explainable deep learning for pulmonary disease and coronavirus COVID-19 detection from X-rays”. In: *Computer Methods and Programs in Biomedicine* 196 (2020), p. 105608.
- [19] Yicheng Fang et al. “Sensitivity of chest CT for COVID-19: comparison to RT-PCR”. In: *Radiology* 296.2 (2020), E115–E117.
- [20] Ophir Gozes et al. “Rapid AI development cycle for the coronavirus (COVID-19) pandemic: Initial results for automated detection patient monitoring using deep learning CT image analysis”. In: *arXiv preprint arXiv:2003.05037* (2020).
- [21] Md Karim et al. “DeepCOVIDExplainer: Explainable COVID-19 predictions based on chest X-ray images”. In: *arXiv preprint arXiv:2004.04582* (2020).
- [22] Asif Iqbal Khan, Junaid Latief Shah, and Mohammad Mudasir Bhat. “CoroNet: A deep neural network for detection and diagnosis of COVID-19 from chest x-ray images”. In: *Computer Methods and Programs in Biomedicine* 196 (2020), p. 105581.

## BIBLIOGRAPHY

- [23] Lin Li et al. “Artificial intelligence distinguishes COVID-19 from community acquired pneumonia on chest CT”. In: *Radiology* (2020).
- [24] Antonios Makris, Ioannis Kontopoulos, and Konstantinos Tserpes. “COVID-19 detection from chest X-Ray images using Deep Learning and Convolutional Neural Networks”. In: *11th Hellenic Conference on Artificial Intelligence*. 2020.
- [25] Abdurrahim Narin, Cankat Kaya, and Zehra Pamuk. “Automatic Detection of Coronavirus Disease (COVID-19) Using X-ray Images and Deep Convolutional Neural Networks”. In: *arXiv preprint arXiv:2003.10849* (2020).
- [26] M.Y. Ng, E.Y. Lee, J. Yang, et al. “Imaging profile of the COVID-19 infection: Radiologic findings and literature review”. In: *Radiology* 2 (2020), e200034.
- [27] Andrea Porzionato et al. “The potential role of the carotid body in COVID-19”. In: *American Journal of Physiology-Lung Cellular and Molecular Physiology* 319.4 (2020), pp. L620–L626.
- [28] Sivaramakrishnan Rajaraman et al. “Iteratively pruned deep learning ensembles for COVID-19 detection in chest X-rays”. In: *IEEE Access* 8 (2020), pp. 115041–115050.
- [29] Fang Song et al. “Emerging 2019 Novel Coronavirus (2019-nCoV) Pneumonia”. In: *Radiology* 295.1 (2020). Erratum in: *Radiology*. 2020 Dec;297(3):E346. PMID: 32027573; PMCID: PMC7233366. doi: 10.1148/radiol.2020200274, pp. 210–217.
- [30] Wenling Wang et al. “Detection of SARS-CoV-2 in different types of clinical specimens”. In: *JAMA* 323.18 (2020), pp. 1843–1844.
- [31] Xingzhi Xie et al. “Chest CT for typical coronavirus disease 2019 (COVID-19) pneumonia: relationship to negative RT-PCR testing”. In: *Radiology* 296.2 (2020), E41–E45.
- [32] Chuansheng Zheng et al. “Deep learning-based detection for COVID-19 from chest CT using weak label”. In: *MedRxiv* (2020).
- [33] Asad Khan, Shahzad Younis, and Haneen Algethami. “COVID-19 Identification Using Deep Neural Networks”. In: *2021 International Conference of Women in Data Science at Taif University (WiDSTaif)*. IEEE. 2021.
- [34] Iqbal H. Sarker. “Machine learning: Algorithms, real-world applications and research directions”. In: *SN Computer Science* 2.3 (2021), pp. 1–21.
- [35] Shuai Wang et al. “A deep learning algorithm using CT images to screen for Corona Virus Disease (COVID-19)”. In: *European Radiology* (2021).

## BIBLIOGRAPHY

- [36] FDA. *FDA Permits Marketing of Artificial Intelligence-Based Device to Detect Certain Diabetes-Related Eye Problems*. Available: <https://www.fda.gov/news-events/press-announcements/fda-permits-marketing-artificial-intelligence-based-device-detect-certain-diabetes-related-eye> (accessed on 15 April 2021).
- [37] Times of India. *Why COVID Testing is a Slow Process and Types of Tests Available*. Available: <https://timesofindia.indiatimes.com/india/why-covid-testing-is-a-slow-process-and-types-of-tests-available/articleshow/76459365.cms>.
- [38] *Vaccinating Machine Learning Against Attacks*. Available: <https://blog.csiro.au/vaccinating-machine-learning-against-attacks/>.
- [39] “World Health Organization (who)”. In: (). Available: <https://www.who.int/emergencies/diseases/novel-coronavirus-2019/situation-reports>.